JOURNAL OF BACTERIOLOGY, June 2005, p. 4023–4032 Vol. 187, No. 12
0021-9193/05/$08.00+0 doi:10.1128/JB.187.12.4023–4032.2005


# Expression Levels Influence Ribosomal Frameshifting at the Tandem Rare Arginine Codons AGG_AGG and AGA_AGA in *Escherichia coli*

Olga L. Gurvich,[1] Pavel V. Baranov,[1,2] Raymond F. Gesteland,[1] and John F. Atkins[1,2]*

*Department of Human Genetics, University of Utah, 15N 2030E, Rm. 7410, Salt Lake City, Utah 84112-5330,[1] and Bioscience Institute, University College Cork, Cork, Ireland[2]*

Received 2 December 2004/Accepted 4 March 2005

**The rare codons AGG and AGA comprise 2% and 4%, respectively, of the arginine codons of *Escherichia coli* K-12, and their cognate tRNAs are sparse. At tandem occurrences of either rare codon, the paucity of cognate aminoacyl tRNAs for the second codon of the pair facilitates peptidyl-tRNA shifting to the +1 frame. However, AGG_AGG and AGA_AGA are not underrepresented and occur 4 and 42 times, respectively, in *E. coli* genes. Searches for corresponding occurrences in other bacteria provide no strong support for the functional utilization of frameshifting at these sequences. All sequences tested in their native context showed 1.5 to 11% frameshifting when expressed from multicopy plasmids. A cassette with one of these sequences singly integrated into the chromosome in stringent cells gave 0.9% frameshifting in contrast to two- to four-times-higher values obtained from multicopy plasmids in stringent cells and eight-times-higher values in relaxed cells. Thus, +1 frameshifting efficiency at AGG_AGG and AGA_AGA is influenced by the mRNA expression level. These tandem rare codons do not occur in highly expressed mRNAs.**

The expression of a minority of genes in probably all organisms utilizes a change in the translational reading frame at a specific site(s). The ribosomal frameshifting involved occurs at shift-prone sequences, and the proportion of ribosomes that participate in frameshifting is greatly elevated in many cases by recoding signals embedded in the mRNA. The "programmed" ribosomal frameshifting in these instances is used for regulatory purposes or to synthesize an additional useful product in a fixed ratio to that of standard decoding (reviewed in references 7, 9, 43, and 56). In some cases, however, only low levels (1 to 2%) of specific frameshifting are important, and the presence, if any, of stimulatory signals aside from the frameshift site and flanking 3′ codon is not evident (10, 23, 67).

Under standard conditions, shift-prone sites exhibit levels of frameshifting orders of magnitude higher than the level of background error frameshifting (1, 35). In addition, certain sequences which exhibit low levels of frameshifting under standard conditions have this level greatly elevated under amino acid starvation (20, 21, 38). Though no examples are known to date, some instances of the latter may be advantageous to the cell in more elaborate ways than simply leading to termination and ribosome recycling. However, under standard conditions, frameshifting at most shift-prone sites, which is not positively utilized for gene expression, likely yields erroneous frameshift products whose synthesis is a waste of energy. Consequently, one might expect shift-prone sequences to be selected against and therefore underrepresented in coding sequences, especially those of highly expressed genes. Conversely, when shift-prone sequences occur, analysis for their possible utilization in gene expression is merited (25, 52).

* Corresponding author. Mailing address: Department of Human Genetics, University of Utah, 15N 2030E, Rm. 7410, Salt Lake City, UT 84112-5330. Phone: (801) 581-5190. Fax: (801) 585-3910. E-mail: john.atkins@genetics.utah.edu.

Here we investigate occurrences of two "shifty" sequences in *Escherichia coli* K-12 genes, determine the level of frameshifting at the sequences, and assess whether the frameshifting at any of these occurrences is likely utilized for gene expression. The two shift-prone sequences are the tandem rare arginine codons AGG_AGG and AGA_AGA. In an earlier investigation which started with overexpression of a mammalian gene in *E. coli*, Spanjaard and van Duin (55) found that translation of AGG_AGG (or AGA_AGA), when present in highly expressed mRNA, yields up to 50% frameshifting. AGG and AGA are decoded by separate (54) and sparse (anticodons $^{3'}$UCC$^{5'}$ and $^{3'}$UCU*$^{5'}$, respectively, where U* is 5-methylaminomethyl-2-thiouridine) tRNAs. Sequestration of the minor tRNA$^{Arg}$ from the limited local pool by the first codon reduces its availability for the second codon of the pair, which then can be considered a "hungry" codon (a term coined by J. Gallant). When the second codon of the pair is at the ribosomal A site, the longer-than-usual time for arrival of its cognate tRNA increases the chance for the dissociation of peptidyl-tRNA, which may re-pair to mRNA at the overlapping +1 frame codon (GGA or GAA). Even though base pairing with the new P site codon is not optimal, the availability of the tRNA cognate to the new A site codon makes frameshifting favorable (6, 46). When this happens, ribosomes continue protein synthesis in the new frame. The introduction of an extra copy of tRNA$^{Arg}_{3'UCC5'}$ or tRNA$^{Arg}_{3'UCU*5'}$ diminishes frameshifting on AGG_AGG and AGA_AGA, respectively (54). This indicates that limitation of the cognate tRNA is responsible for the "shiftiness" of the AGG_AGG and AGA_AGA sequences. Since separate tRNAs decode AGG and AGA in *E. coli*, it is not surprising that the mixed codon pairs AGG_AGA and AGA_AGG exhibit undetectable or greatly reduced frameshifting (54).

Previous studies have shown the importance of tRNA bal-

TABLE 1. *E. coli* strains and plasmids used in this study

| Strain or plasmid | Genotype and/or characteristic(s) | Reference or source | Experiment used |
|---|---|---|---|
| Parental strains | | | |
| DH5α | *endA1 recA1 relA1 gyrA96 hsdR17* ($r_K^-$ $m_K^+$) *phoA supE44 thi-1* Δ(*lacZYA-argF*)*U169* Φ*80* Δ(*lacZ*)M15 $F^-$ | Lab stock | Pulse-chase |
| CP78 MC1000 | *thr leu his arg thi* Δ*lac* | 17 (obtained from J. Gallant) | Nonintegrated β-Gal assay |
| CP79 MC1000 | *thr leu his arg thi* Δ*lac relA2* | 17 (obtained from J. Gallant) | Nonintegrated β-Gal assay |
| BW27786 | *lacI*$^q$ *rrnB3* Δ*lacZ4787 hsdR514* DE(*araBAD*)*567* DE(*rhaBAD*)*568* DE(*araFGH*) Φ(Δ*araEp* Pcp13-*araE*) | 34 (obtained from *E. coli* Genetic Stock Center) | |
| New strains | | | |
| JFG001 | BW27786 $P_{araB}$-GST-*emrK-lacZ*(+1), *lacZ* in +1 frame relative to GST | This study | Integrated β-Gal assay |
| JFG002 | BW27786 $P_{araB}$-GST-*emrK-lacZ*(0), *lacZ* in 0 frame relative to GST | This study | Integrated β-Gal assay |
| Parental vectors | | | |
| pGHM57 | $P_{tac}$ GST-MBP pBR322 ori *bla lacI*$^q$ | 29 | |
| pSKAGS | $P_{tac}$ GST-*lacZ* pBR322 ori *bla* | 66 | |
| pLacI$^q$ | *lacI*$^q$ *kan* | Gift from R. Maldonado | |
| pLA2 | $P_{araB}$ *lacZ kan*, λ integration site, *oriR* | 26 (obtained from *E. coli* Genetic Stock Center) | |
| pINT-ts | λ integrase | 26 (obtained from *E. coli* Genetic Stock Center) | |
| New vectors | | | |
| pGHM-*yhaC* | pGHM57 GST-*yhaC*-MBP | This study | Pulse-chase |
| pGHM-*smf* | pGHM57 GST-*smf*-MBP | This study | Pulse-chase |
| pGHM-*fecC* | pGHM57 GST-*fecC*-MBP | This study | Pulse-chase |
| pGHM-*lhr* | pGHM57 GST-*thr*-MBP | This study | Pulse-chase |
| pGHM-*recF* | pGHM57 GST-*recF*-MBP | This study | Pulse-chase |
| pGHM-*t150* | pGHM57 GST-*t150*-MBP | This study | Pulse-chase |
| pGHM-*emrK* | pGHM57GST-*emrK*-MBP | This study | Pulse-chase |
| pSKAGS-*emrK*+1 | pSKAGS GST-*emrK-lacZ*, *lacZ* in +1 frame relative to GST | This study | Nonintegrated β-Gal assay |
| pSKAGS-*emrK*-0 | pSKAGS GST-*emrK-lacZ*, *lacZ* in 0 frame relative to GST | This study | Nonintegrated β-Gal assay |
| pLA2-GST-*emrK*+1 | pLA2 GST-*emrK-lacZ*, *lacZ* in +1 frame relative to GST | This study | Integrated β-Gal assay |
| pLA2-GST-*emrK*-0 | pLA2 GST-*emrK-lacZ*, *lacZ* in 0 frame relative to GST | This study | Integrated β-Gal assay |

ance for frameshifting (2, 21, 45, 58). A corollary is that the abundance of mRNAs containing sequences prone to frameshifting due to tRNA limitation is expected to be equally important. Such mRNAs, when expressed at high levels, can sequester a significant fraction of the rare tRNAs at the first of the tandem codons, thereby significantly depleting the pool of free tRNAs and increasing the frequency of frameshifting. Spanjaard and van Duin (55) noted an unspecified decrease in frameshifting when the mRNA concentration was reduced, although no data were presented. Here, the efficiency of frameshifting in the native-gene context was also examined under different mRNA expression levels.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** The *E. coli* strains and plasmids used in this study are listed in Table 1. Gene sequences (*yhaC*, *smf*, *fecC*, *lhr*, *recF*, *t150*, and *emrK*) were amplified by PCR from *E. coli* genomic DNA by use of primers with appropriate overhanging restriction sites and cloned into the BamHI/EcoRI sites in between glutathione *S*-transferase (GST) and maltose binding protein (MBP) of pGHM57, so that MBP was in the +1 frame relative to GST. Additionally, the *emrK* sequence (coding for multidrug efflux protein) was cloned into the HindIII/ApaI sites in between GST and *lacZ* of pSKAGS, so that *lacZ* was either in the +1 frame relative to GST (frameshift reporter) or in the 0 frame (in-frame control). Further, the GST-*emrK* fusions were amplified by PCR from the constructions in the pSKAGS vector and cloned into the NdeI site of pLA2 in between $P_{araB}$ and *lacZ*. All plasmid constructions were confirmed by DNA sequencing on automated sequencing machines (model ABI3730).

**Plasmid integration.** The JFG001 and JFG002 integrant strains were constructed as described by Haldimann and Wanner (26). Briefly, electrocompetent BW27786 cells harboring pINT-ts helper plasmid were transformed with pLA2(GST-*emrK*) vectors, incubated for 1 h at 37°C and for 30 min at 42°C, spread onto agar containing 10 μg/ml kanamycin, and incubated at 37°C. Colonies were tested by PCR for copy number according to the method of Haldiman and Wanner (26). Single-copy integrants were then purified once nonselectively, tested for antibiotic resistance for stable integration and loss of the helper plasmid, and retested for copy number.

**Estimate of frameshift efficiency by pulse-chase analysis.** Overnight cultures of strains expressing the appropriate construct were grown in MOPS (morpho-

linepropanesulfonic acid)-glucose (44) containing 100 μg/ml ampicillin and all amino acids (150 μg/ml each) except methionine and tyrosine and diluted 1:50 in 300 μl of the same medium. After 2 h of incubation at 37°C, cultures were induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 10 min. The cells were pulse-labeled for 2 min by the addition of 7.5 μCi [$^{35}$S]methionine in 30 μl of medium, chased for 2 min by the addition of 30 μl of 50-mg/ml cold methionine, chilled on ice, and harvested by centrifugation. The pellets were resuspended in 50 μl cracking buffer (6 M urea, 1% sodium dodecyl sulfate, 50 mM Tris-HCl [pH 7.2]) and heated at 95°C for 5 min. Five-microliter aliquots were loaded on 4 to 12% NuPAGE gels (Invitrogen, Inc.) and electrophoresed, under conditions recommended by the manufacturer, in MOPS-sodium dodecyl sulfate buffer (Invitrogen, Inc.). Gels were exposed overnight and visualized with a Molecular Dynamics PhosphorImager. The amounts of termination and frameshift products were quantified by ImageQuant. The frameshifting efficiency was estimated as the ratio of the amount of frameshift product to the sum of the termination and frameshift products.

**Estimate of frameshift efficiency by β-galactosidase assays.** Overnight cultures of CP78/CP79 harboring the pSKAGS-*emrK* constructs and pLacI$^q$ or JFG001 and JFG002 integrant strains were grown in MOPS-glucose (44) containing appropriate antibiotics and all amino acids (150 μg/ml each) except tyrosine and diluted 1:50 in 800 μl of the same medium. After 2 h of incubation at 37°C, half of the cultures were induced (with either 1 mM IPTG or 0.02% or 0.2% L-arabinose), and incubation continued for another hour. Cultures were chilled on ice, and β-galactosidase assays were performed as described by Miller (42). Assays were performed in triplicate on three separate days with at least two independent clones/integrants. The frameshifting efficiency was calculated as the ratio of β-galactosidase activity of the cells harboring the frameshift reporter to the β-galactosidase activity of the cells harboring the in-frame control construct.

**Random-genome analysis.** The CodonShuffle and the DicodonShuffle programs were obtained from Christopher Burge (33). The NC_000913.ffn file (the 24 June 2004 updated version is gi 49175990), which contains the nucleotide sequences of all *E. coli* K-12 protein-coding genes, was downloaded from the National Center for Biotechnology Information's website (http://www.ncbi.nlm.nih.gov/). The MotifCounter program was written with Java2 (http://java.sun.com/) and BioJava (http://www.biojava.org/) (25).

The DicodonShuffle or the CodonShuffle program was used to generate 1,000 randomized nucleotide files from the original NC_000913.ffn file. The MotifCounter was then used to search through the 1,000 new files and to count how many times a motif occurred in each genome. The results were further analyzed in Microsoft Excel for the distribution of a motif in 1,000 random genomes.

**Calculation of the total number of genes in which error frameshifting occurs.** AGA_AGA occurs 42 times in the genome; however, in three genes, it occurs twice. Therefore, there are 39 AGA_AGA-containing genes. AGG_AGG occurs four times in the genome in four different genes. CCC_TGA occurs 19 times in the *E. coli* K-12 genome; however, frameshifting was experimentally verified to occur in only nine genes. A_AAA_AAG occurs 70 times in *E. coli* K-12: 9 times in programmed frameshifting cases, 2 times in non-protein-encoding genes (misannotated; see reference 25), and 1 time in a gene that also contains CCC_TGA and therefore has been taken into account already. We presume that all genes that contain AGG_AGG, AGA_AGA, and A_AAA_AAG exhibit some level of frameshifting, since frameshifting was experimentally detected in all of the genes tested that contain these sequences. Therefore, the total number of genes in which error frameshifting occurs is 108 (39 AGA_AGA plus 4 AGG_AGG plus 9 CCC_TGA plus 56 A_AAA_AAG). There are 4,242 protein-encoding genes annotated in *E. coli* K-12 (NC_000913.ffn; gi 49175990), of which 108 genes constitute about 2.5%.

## RESULTS

**Occurrence of tandem AGG and AGA codons in *E. coli* K-12 genes.** A search for AGG_AGG and AGA_AGA sequences in annotated *E. coli* K-12 open reading frames (ORFs) revealed that there are 4 occurrences of AGG_AGG and 44 occurrences of AGA_AGA (genes containing these sequences are listed in Table 2). However, two of the genes containing AGA_AGA, *yjbF* and *pnp*, were excluded from the final count of the AGA_AGA occurrences and from the analysis which follows. For both genes, the predicted start site is not conserved among different species and lacks an appropriate Shine-Dalgarno sequence upstream. The probable initiation site is located downstream of AGA_AGA. In fact, a downstream start site was previously reported for *pnp* (47). Both genes were annotated so that they contained AGA_AGA in the *E. coli* K-12 genome version available in August 2003 when the work was initiated and in the final version from 24 June 2004 (gi 49175990) used in the text. However, several intermediate releases (i.e., gi 16127994) had start sites annotated downstream of AGA_AGA for both genes. Therefore, we presume that these genes do not contain AGA_AGA and that the total count of AGA_AGA is 42. (Two more genes, *b2655* and *livJ*, were annotated to contain AGA_AGA in their coding sequences in the August 2003 genome version. Alignment of the *b2655* sequences from different bacteria showed that there is a nucleotide insertion upstream of the AGAAGA sequence in the *E. coli* K-12 gene. In other bacteria, AGAAGA in *b2655* is read in the frame NNA_GAA_GAN. When this portion of the gene was amplified by PCR from the DH5α strain [a derivative of K-12] and sequenced, we found that the frameshift mutation was not present and that AGAAGA was read in the same frame as in other bacteria. The June 2004 release of the *E. coli* K-12 genome has the corrected sequence. Our analysis of *livJ* suggested that the start codon is located downstream of AGA_AGA; a subsequent genome version also had the start codon annotated downstream of AGA_AGA. Consequently, it is not surprising that Boycheva et al. [12] reported 45 occurrences of AGA_AGA in the *E. coli* K-12 genome, since different releases do vary slightly.)

A statistical analysis was performed to find out whether AGG_AGG and AGA_AGA are over- or underrepresented in coding regions. For this purpose, the DicodonShuffle program was employed (33). The DicodonShuffle program generates random genomes from the real genome in question (i.e., *E. coli* K-12) by randomizing mRNA sequences while preserving the sequences of the encoded protein, the codon usage, and the dinucleotide composition of the original message. Consequently, the random genome maintains the same amino acid constraints as the real genome but is relieved from the nucleotide constraints. One thousand random genomes were generated, and the distribution of sequences in question was analyzed (Fig. 1). The mean values of AGG_AGG and AGA_AGA occurrences in the 1,000 generated genomes are 5.1 and 38.3, respectively. The observed numbers in real genomes of 4 for AGG_AGG and 42 for AGA_AGA are within the distribution range and are anticipated in the absence of selective pressure (Fig. 1A). It should be noted that the value 38.3 predicted here by the DicodonShuffle program for the AGA_AGA occurrences is strikingly different from the expected value of 6.2 reported by Boycheva et al. (12). However, in their estimate, they took into account only the codon usage bias of the native genome. When we generated the random genomes using the CodonShuffle program (33), which does not take into account the dinucleotide composition and is similar to an approach used before (25, 51), the AGA_AGA occurrence was predicted to be 7.2. Thus, apparently the dinucleotide composition of the *E. coli* K-12 genome contributes greatly to the occurrence of AGA_AGA.

**Distribution analysis of AGGAGG in other frames.** AGGAGG is a strong Shine-Dalgarno sequence, in contrast to AGAAGA, and can functionally interact with translating ribosomes in a frame-independent manner (see below). Therefore,

TABLE 2. *E. coli* K-12 genes containing AGG_AGG or AGA_AGA sequences

| Gene name | Protein function | Gene length (nt) | Positions of tandem codons (nt) | No. of sense codons after +1 shift | Conservation of tandem codon[a] | Protein conservation[a] |
|---|---|---|---|---|---|---|
| AGG_AGG-containing genes | | | | | | |
| *ninE* | Unknown | 171 | 154, 160 | 113 | K-12, O157, phage 82 | K-12, O157, bacteriophages |
| *yhaC* | Unknown | 1,188 | 409, 415 | 1 | E.c., S.f. | E.c., S.f. |
| *smf* | Putative Rossmann-fold nucleotide-binding protein | 1,125 | 1084, 1090 | 471 | E.c., S.f., S.t., Y.p. | Conserved |
| *fecC* | Citrate-dependent iron(III) transport | 999 | 985, 991 | 90 | K-12 | Conserved |
| AGA_AGA-containing genes | | | | | | |
| *intE* | Prophage e14 integrase | 1,128 | 556, 561 | 7 | K-12 | Conserved |
| *rfbX* | Putative O-antigen transporter | 1,248 | 25, 30 | 24 | K-12 | Conserved |
| *yagM* | CP4-6 prophage | 855 | 685, 690 | 45 | K-12 | Conserved |
| *yacH* | Putative membrane protein | 1,854 | 1837, 1842 | 8 | K-12 | Conserved |
| *ydaU* | Rac prophage | 858 | 370, 375 | 15 | K-12 | K-12 |
| *ymfK* | e14 prophage, putative phage repressor | 675 | 652, 657 | 10 | K-12 | K-12 |
| *rfaS* | Lipopolysaccharide core biosynthesis | 936 | 640, 645 | 2 | K-12 | K-12 |
| *ymfH* | e14 prophage | 312 | 196, 201 | 7 | K-12 | K-12 |
| *ydfO* | Qin prophage | 426 | 289, 294 | 46 | K-12, CFT073 | E.c. |
| *gspA* | Putative export protein A | 1,470 | 10, 15 | 40 | K-12, CFT073 | Conserved |
| *ycdF* | Unknown | 231 | 199, 204 | 44 | K-12, CFT073, S.f. | K-12, CFT073, S.f. |
| *ygeP* | Unknown | 300 | 280, 285 | 5 | K-12, O157 | K-12, O157 |
| *yjcF* | Unknown | 1,293 | 325, 330 | 17 | K-12, O157, S.f. | K-12, O157, S.f. |
| *yhiI* | Unknown | 1,623 | 799, 804 | 4 | K-12, O157, S.f. | K-12, O157, S.f. |
| *t150* | IS*150* putative transposase | 852 | 586, 591 | 33 | K-12, S.f. | Conserved |
| *t150* | IS*150* putative transposase | 852 | 778, 783 | 3 | K-12, S.f. | Conserved |
| *ylbH* | Unknown | 711 | 307, 312 | 5 | E.c. | E.c. |
| *b1459* | Unknown | 201 | 181, 186 | 5 | E.c. | Conserved |
| *ydeN* | Putative sulfatase | 1,716 | 19, 24 | 4 | E.c. | Conserved |
| *emrK* | Multidrug resistance protein K | 1,164 | 37, 42 | 8 | E.c. | Conserved |
| *intC* | Putative prophage Sf6-like integrase | 1,158 | 484, 489 | 26 | E.c. | Conserved |
| *sfmF* | Putative fimbria-like protein | 516 | 4, 9 | 15 | E.c. | Conserved |
| *yfcC* | Putative *S*-transferase | 1,542 | 52, 57 | 8 | E.c. | Conserved |
| *ygeH* | Putative invasion protein | 1,377 | 1210, 1215 | 5 | E.c. | Conserved |
| *yhiU* | Putative membrane protein | 1,158 | 7, 13 + 11, 16 | 3 or 4 | E.c. | Conserved |
| *ybcK* | DLP12 prophage, putative recombinase | 1,527 | 73, 78 | 5 | E.c. | Conserved |
| *ybcK* | DLP12 prophage, putative recombinase | 1,527 | 1000, 1005 | 8 | E.c. | Conserved |
| *ybfL* | Putative receptor | 858 | 694, 699 | 17 | E.c. | Conserved |
| *ydcC* | H repeat-associated protein | 1,137 | 973, 978 | 17 | E.c. | Conserved |
| *yhhI* | H repeat-associated protein | 1,137 | 973, 978 | 17 | E.c. | Conserved |
| *yjgR* | Putative nucleotide triphosphate hydrolase | 1,503 | 1492, 1497 | 60 | E.c. | Conserved |
| *yqeI* | Putative sensory transducer | 810 | 292, 297 | 14 | E.c. | Conserved |
| *intR* | Rac prophage, putative transposase | 1,236 | 757, 7632 | 2 | E.c., S.f. | Conserved |
| *ynbB* | Putative phosphatidate cytidiltransferase | 897 | 577, 582 | 15 | E.c., S.f. | Conserved |
| *lhr* | Enzyme; DNA replication and repair | 4,617 | 2797, 2802 | 15 | E.c., S.f. | Conserved |

*Continued on following page*

TABLE 2—*Continued*

| Gene name | Protein function | Gene length (nt) | Positions of tandem codons (nt) | No. of sense codons after +1 shift | Conservation of tandem codon[a] | Protein conservation[a] |
|---|---|---|---|---|---|---|
| *yddW* | Unknown | 1,320 | 37, 42 | 4 | E.c., S.f. | Conserved |
| *yifQ* | Unknown | 726 | 703, 708 | 12 | E.c., S.f. | Conserved |
| *mdoC* | Membrane protein for succinylation of osmoregulated periplasmic glucans | 1,158 | 580, 585 | 10 | E.c., S.f. | Conserved |
| *ybjR* | *N*-Acetylmuramoyl-L-alanine amidase | 831 | 4, 9 | 45 | E.c., S.f. | Conserved |
| *recF* | Gap repair protein | 1,074 | 394, 399 | 35 | E.c., S.f., S.t. | Conserved |
| *trmD* | tRNA (guanine-1-)-methyltransferase | 768 | 655, 660 | 9 | E.c., S.f., S.t., Y.p. | Conserved |
| *yjbF*[b] | Putative membrane-associated protein | 669 | 22, 27 | 2 | E.c., S.f. | Conserved |
| *pnp*[b] | Polynucleotide phosphorylase | 2,205 | 7, 12 | 7 | E.c., S.f. | Conserved |

[a] Bacterial names are abbreviated as follows: K-12, *E. coli* K-12; O157, *E. coli* O157:H7; CFT073, *E. coli* CFT073; E.c. includes all three *E. coli* species (K-12, O157:H7, and CFT073); S.f., *Shigella flexneri* 2a strain 301; S.t., *Salmonella enterica* serovar Typhimurium LT2; Y.p., *Yersinia pestis* KIM. A protein is considered conserved if its homologues are present in at least four different genera.

[b] The genes *yjbF* and *pnp* were excluded from the total count of the AGA_AGA occurrences and all other considerations in the text for the reasons discussed in the text.

its distribution was examined in other translational frames in *E. coli* K-12 genes to determine whether its representation is biased. The sequence NAG_GAG_GNN occurs 168 times in *E. coli* genes. In 1,000 random genomes generated by the DicodonShuffle program, the sequence NAG_GAG_GNN occurs 396 times per genome on average, and the lowest count is 342 times in one genome (Fig. 1C). The sequence NNA_GGA_GGN occurs 47 times in coding regions of *E. coli* K-12. In

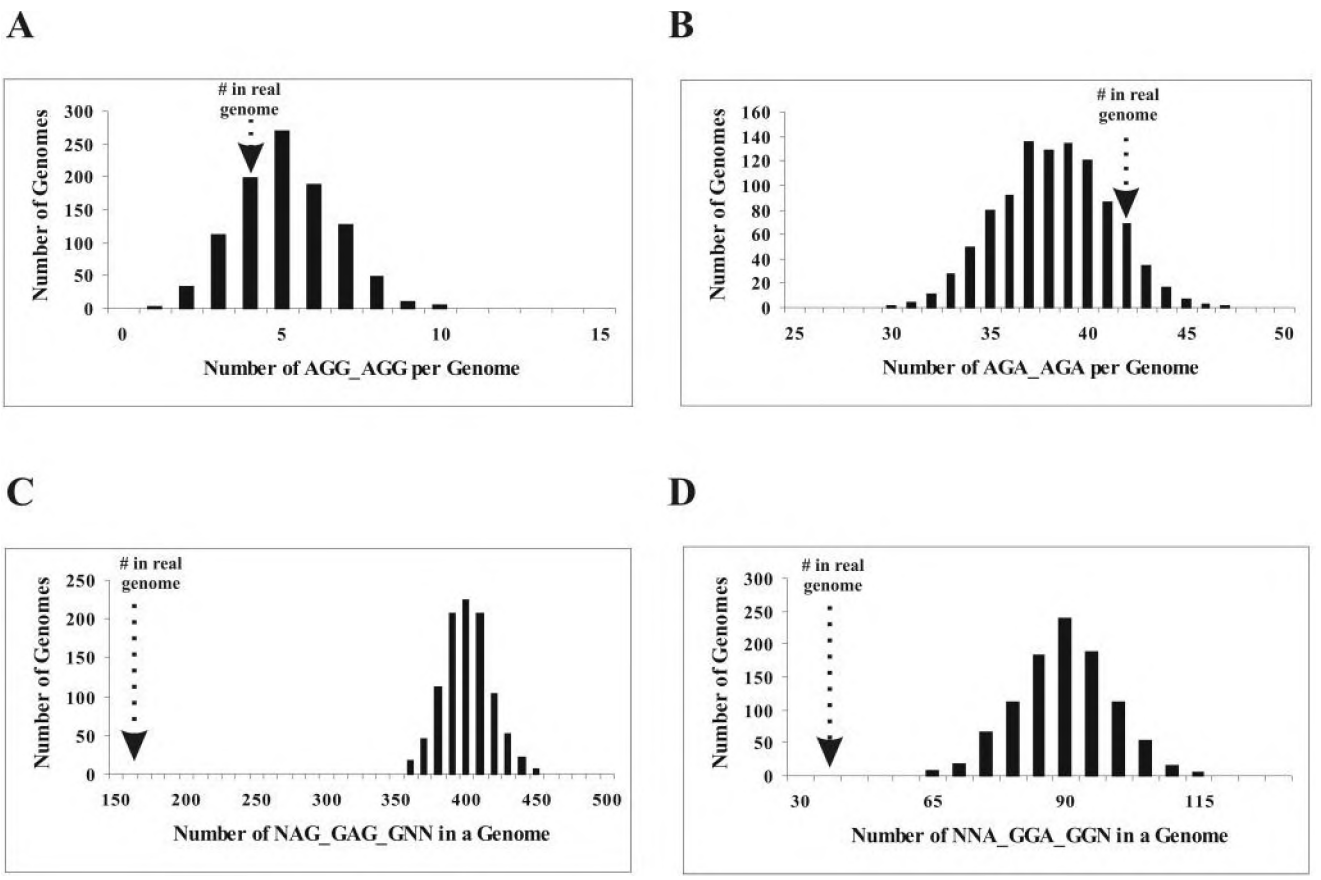


FIG. 1. Distribution of occurrences of AGGAGG and AGAAGA sequences in 1,000 computer-generated random genomes.

TABLE 3. Frameshifting efficiencies on AGA_AGA in an *emrK* context under different expression conditions

| Plasmid/strain | Reporter copy no. | Promoter | Inducer | β-Galactosidase activity (Miller units)[a] | | Average frameshifting level (%) ± SD |
|---|---|---|---|---|---|---|
| | | | | Frameshift reporter | In-frame control | |
| pGHM-*emrK*/DH5α | 15–20 | $P_{tac}$ | 1 mM IPTG | NA | NA | 7.4 ± 3.4 |
| pSKAGS-*emrK*/CP79 | 15–20 | $P_{tac}$ | None | 41 ± 2 | 1,847 ± 105 | 2.2 ± 0.2 |
| | | | 1 mM IPTG | 376 ± 33 | 4,698 ± 545 | 8.0 ± 1.2 |
| pSKAGS-*emrK*/CP78 | 15–20 | $P_{tac}$ | None | 10 ± 1 | 690 ± 67 | 1.5 ± 0.1 |
| | | | 1 mM IPTG | 63 ± 6 | 4,559 ± 348 | 1.4 ± 0.2 |
| pSKAGS-*emrK*/BW27786 | 15–20 | $P_{tac}$ | None | 20 ± 2 | 1,449 ± 160 | 1.4 ± 0.2 |
| | | | 1 mM IPTG | 345 ± 55 | 8,850 ± 465 | 3.9 ± 0.7 |
| JFG001 (derived from BW27786) | 1 | $P_{araB}$ | 0.02% arabinose | 1.6 ± 0.1 | 165 ± 15 | 1.0 ± 0.1 |
| | | | 0.2% arabinose | 2.6 ± 0.3 | 287 ± 13 | 0.9 ± 0.1 |

[a] NA, not applicable.

1,000 random genomes, the same sequence occurs 88 times per genome on average, and the lowest count is 61 times in a genome (Fig. 1D). Thus, the AGGAGG sequence is significantly underrepresented in the other two frames.

**Comparative analysis of AGG_AGG- and AGA_AGA-containing genes.** In genes that exploit programmed ribosomal frameshifting for expression purposes, the shift site is often conserved (the +1 frameshifting utilized in decoding many eubacterial RF2 genes is an example [8]). Where the ORF in the new frame after the frameshift site is long, the sequence of the product from that frame is also conserved (for example, in RF2 and antizyme [8, 30, 41]). (Exceptions are candidates for hypothetical nonproduct roles for frameshifting [3]). Such conservation makes it easier to list viable candidates for this type of programmed frameshifting than when the ribosome quickly terminates and the purpose is to generate a product lacking significant domains present in the full-length product of standard decoding (e.g., the gamma subunit of DNA polymerase III encoded by *E. coli dnaX* [11, 18, 61]). Of course, as also illustrated by *dnaX* decoding, frameshifting may be important for gene expression but only somewhat conserved. In some bacteria, the same product mix and ratio are achieved by a different mechanism, e.g., transcription slippage (37).

All *E. coli* K-12 genes that contain tandem AGG or AGA codons were compared with homologous genes in other bacteria for conservation of the frameshift site and the amino acid sequence encoded in the +1 frame after the shift site. However, in none of the genes containing AGG_AGG or AGA_AGA can the utilization of frameshifting for gene expression purposes be predicted with high certainty.

**Efficiency of frameshifting on AGG_AGG and AGA_AGA in their native contexts.** As described in the introduction, the efficiency of frameshifting on AGG_AGG and AGA_AGA tandem codons at different expression levels merits study. We first attempted to achieve the lowest practical expression levels. For that, the *lacZ* reporters with the AGA_AGA-containing sequence of the multidrug efflux gene *emrK* were integrated into the *E. coli* chromosome. The frameshifting efficiency was assayed by β-galactosidase assays. In this setting, frameshifting is detectable at about 0.9 to 1% (Table 3). Thus, the translation of an AGA_AGA-containing mRNA synthesized from a single-copy gene from the $P_{araB}$ promoter still results in a detectable level of frameshifting, albeit about 50 times lower than that reported previously (55). Although it is possible to titrate the level of inducer (L-arabinose) and achieve lower levels of mRNA expression, the data would be difficult to reliably quantitate since, in the performed experiment, β-galactosidase activity is rather low even for the in-frame control (165 and 287 Miller units with 0.02% and 0.2% arabinose induction, respectively).

Next, the levels of frameshifting were assayed when the same sequences were expressed from pSKAGS-based plasmid-borne constructs (with pBR322 ori). The expression of a GST-*emrK-lacZ* fusion in this vector is driven by a $P_{tac}$ promoter and, since the promoter is leaky, can be measured with and without IPTG induction (approximately a fivefold difference in the expression levels judging from expression levels from the in-frame controls). The β-galactosidase assays were then performed with the same strain as above. In the absence of IPTG induction, the frameshifting level is 1.4%, while in the presence of inducer, the frameshift level is 3.9%. The respective values for the in-frame controls are 1,449 and 8,850 Miller units (Table 3).

With the same pSKAGS-*emrK* constructs, frameshifting efficiency was also measured in *relA*$^+$ and *relA* mutant isogenic strains (CP78 and CP79), since starvation-promoted frameshifting is influenced by *relA* gene inactivation (19, 39). In the absence of IPTG induction, frameshifting is 1.5% and 2.2% in these *relA*$^+$ and *relA* mutant strains, respectively (Table 3). In the presence of the IPTG inducer, frameshifting levels on the AGA_AGA-containing sequence in the *relA*$^+$ strain did not show any significant change. On induction in the *relA* mutant strain, however, frameshifting increased to 8%. A similar value of frameshifting, 7.4%, was obtained for this sequence when frameshifting was measured by pulse-chase assay in the DH5α strain (also a *relA* mutant) in the presence of the inducer.

The pulse-chase method is simpler and more reliable, since both products, of termination and frameshifting, can be visualized. Therefore, we employed pulse-chase analysis to examine frameshifting in cassettes with the sequence context from

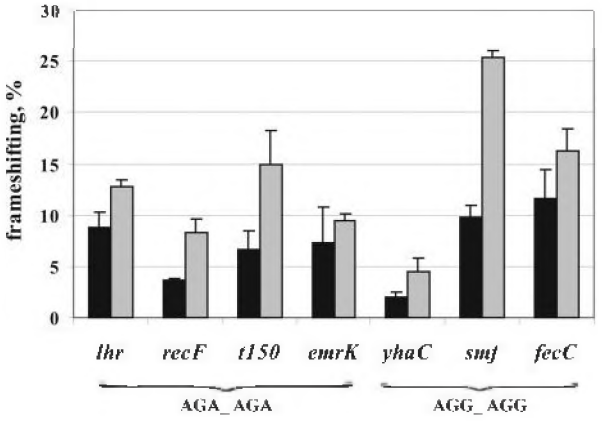


FIG. 2. Frameshifting efficiencies on AGG_AGG and AGA_AGA tandem codons in *E. coli* K-12 genes at 37°C (black bars) and at 42°C (grey bars).

several genes in which the tandem rare codons naturally occur. Frameshifting was measured from cassettes with the sequence contexts from three AGG_AGG-containing (*yhaC*, *smf*, and *fecC*) and three other AGA_AGA-containing (*lhr*, *recF*, and *t150*) genes. Although tentatively proposed only in one case of prokaryotic +1 frameshifting (69), 3′ mRNA structures might stimulate +1 frameshifting (and are known for eukaryotic antizyme +1 frameshifting [30, 41]). Therefore, all sequences were examined for the presence of potential stem-loop structures downstream of the shift site. When no strong candidate was identified, sequences from the genes encompassing at least 30 nucleotides upstream and downstream of the shift site (or up to, but excluding, the stop codon in the +1 frame) were included in the constructs. Pulse-chase experiments were performed at 37°C. Frameshifting was detected in all genes examined. The efficiency of frameshifting ranges from 3 to 11% (Fig. 2), which is much lower than that reported before, though the prior analysis was performed at 42°C (55). Since frameshifting efficiencies can be up to fivefold higher at 42°C than at 37°C on some sequences (O. Gurvich, unpublished data), the pulse-chase analysis was redone at 42°C. With the AGG_AGG and AGA_AGA sequences in their native flanking sequence contexts, the frameshifting values obtained at 42°C are up to two times higher than those at 37°C under the conditions tested (Fig. 2). However, they are still much lower than the 50% level reported previously. We were unable to discern from the original Spanjaard and van Duin paper (55) the flanking sequence, mainly from reporter genes, used in that work and so were unable to include it for comparative purposes in the present experiments. There may well be avoidance in the natural gene context of features that would stimulate high levels of erroneous frameshifting at tandem AGG and tandem AGA codons, which are fortuitously present in other sequences.

## DISCUSSION

As foreshadowed in the study of Spanjaard and van Duin (55), and investigated here, the level of frameshifting at the tandem rare codon AGA_AGA (and presumably at AGG_AGG) is influenced by the codon's expression level. When expression was tested in the same strain (BW27786), there was a fourfold decrease in frameshifting when AGA_AGA was integrated in the chromosome and expressed from a $P_{araB}$ promoter than when it was borne on a pBR322-based plasmid and expressed from an induced *tac* promoter. The relative expression, as quantified by comparing β-galactosidase activities from the in-frame controls, is 53-fold lower from the integrated reporter induced with 0.02% arabinose than from the plasmid-borne cassette induced with 1 mM IPTG. The higher expression of the AGA_AGA-containing mRNA likely results in the sequestration of the sparse cognate tRNA into separate translating complexes. Thus, more ribosomes translating AGA_AGA are paused at the first AGA codon, and more of them can escape stalling by shifting into the +1 frame. These results are in good agreement with a previous report (68) that overexpression of mRNA enriched with AGA and AGG codons results in a translation inhibition which is reversible by increasing the levels of the rare $tRNA^{Arg}$.

As shown above, the frameshifting is exacerbated upon induction in the *relA* mutant, rather than in the otherwise identical *relA*$^+$, strain. Masucci et al. (39) proposed that starvation-promoted frameshifting is higher in *relA* mutant cells than in *relA*$^+$ cells due to accumulation of undermodified tRNA during amino acid starvation. Several tRNA modifications prevent +1 frameshifting (62), and the accumulation of the hypomodified tRNAs in *relA* mutant cells is the likely cause of the higher frameshifting under relaxed conditions. The above-described experiments were not performed under amino acid starvation per se. However, the overexpression of mRNA containing AGA_AGA increases the demand for the cognate tRNA dramatically and is therefore likely to trigger a response similar to that of starvation. The *relA*$^+$ strain is better able to adjust the levels of cognate tRNA and so to keep frameshifting levels lower than those of the *relA* mutant cells.

All gene sequences tested supported +1 frameshifting with efficiencies of 3 to 11%. Nevertheless, this plasmid-borne frameshifting is still at least 4.5-fold less than the 50% value reported by Spanjaard and van Duin (55). Their study also used a pBR322-based plasmid. However, there are several discrepancies which together likely account for the different values obtained. Spanjaard and van Duin used a $P_L$ promoter (whose induction required 42°C thermal inactivation of a thermolabile repressor; the elevated temperature itself may have modestly elevated the frameshifting level compared to 37°C). They also had an unknown sequence context which may have contained additional rare arginine codons which would affect the cognate tRNA pool levels. Moreover, although Spanjaard and van Duin carried out their experiments with two completely different strains, both strains carry mutations conferring streptomycin resistance. Certain streptomycin-resistant ribosomes translate more accurately and more slowly (50). Such hyperaccurate ribosomes are expected to spend more time with their A sites vacant. As shown by Sipley and Goldman (53), such ribosomes are more prone to shift into the +1 frame (though this does not imply that the leakiness of all frameshift mutants will be elevated in the streptomycin-resistant strains).

None of the *E. coli* K-12 chromosomal genes that contain AGG_AGG or AGA_AGA belong to the subset of highly expressed genes. Low expression could be due to any of several causes, e.g., low promoter strength or low translation rate due to suboptimal codon usage. The latter situation would be sim-

ilar to that of the AGA_AGA-containing *trmD* gene, which is transcribed at a high level but is translated poorly and expressed only at about 80 protein molecules per cell (13). Some of the AGG_AGG- or AGA_AGA-containing genes are most likely not expressed under standard conditions but become activated depending on growth medium and physiological state. This is certainly true for *emrK*, the drug efflux gene that we used for comparison studies, the transcription of which is dependent on the drug presence and growth phase (16, 40, 60). Thus, at a given time, most likely only a subset of these genes is expressed, and the total amount of mRNAs containing AG A_AGA (or AGG_AGG) is likely to be somewhat lower than that from the $P_{araB}$ promoter used in the current study (at 0.02% arabinose induction). Consequently, the frameshifting efficiencies exhibited during expression of these genes in *E. coli* are presumably lower. However, the natural frameshifting levels are also likely to vary depending on growth condition and physiological state. When *E. coli* enters the stationary phase, frameshifting levels increase (5, 65). In contrast, *Saccharomyces cerevisiae* Ty1-programmed frameshifting decreases on entering the stationary phase, probably due to a disproportionate decrease in overall demand for the sparse tRNA that mediates the shift (57).

The most likely efficiencies of frameshifting at AGA_AGA and AGG_AGG in the mRNAs in which they naturally occur are at least 1 to 2 orders of magnitude higher than the level estimated for general coding sequences (35). Even so, any results of selection against occurrences of these sequences in genes that are not highly expressed are not evident. AGG _AGG is predicted to occur five times, but in reality, it occurs four times in coding regions of *E. coli* K-12. In three of these occurrences, in *fecC*, *smf*, and *ninE*, the AGG_AGG sequence likely serves as a Shine-Dalgarno sequence for the initiation of translation of the downstream gene. The benefit of this sequence for translation initiation likely outweighs any negative consequences of its frameshifting propensity, probably because the products of frameshifting are not toxic and/or rapidly degraded. In the fourth gene, *yhaC*, there is no obvious Shine-Dalgarno role for the AGG_AGG sequence. It could be that this sequence is involved in the formation of some RNA secondary structure and therefore is selected for. Indeed, when the nucleotide sequence of this part of the gene was analyzed by mfold (70), a potential strong stem-loop structure with an AGGAGG sequence involved in base pairing was predicted. However, it is hard to assess by covariation analysis whether the stem-loop is functional, since this gene exists only in *E. coli* and *Shigella flexneri* and the nucleotide sequences of these genes are nearly identical.

**Pairing of AGGAGG with rRNA likely influences its occurrence within coding sequences.** AGGAGG is one of the strongest Shine-Dalgarno sequences and, by virtue of its binding to 16S rRNA in 30S ribosomal subunits, it directs translation initiation. However, when it occurs in coding sequences, it is also sensed by translating 70S ribosomes, due to the scanning of mRNA by the anti-Shine-Dalgarno sequence in these ribosomes (63, 64). This can result in mRNA-rRNA pairing, which can potentially stall ribosomes. The pushing or pulling effect of the internal Shine-Dalgarno sequence in these instances, with the effect and directionality depending on the spacing length from the shift site (4, 36), might be expected to result in selective pressure against occurrences of AGGAGG when frameshifting is not functionally favored, though evidence for this has not been forthcoming.

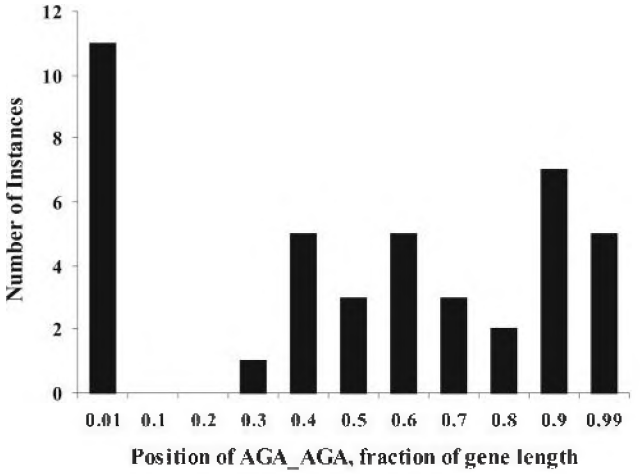


FIG. 3. Positions of AGA_AGA tandem codons normalized to the gene length.

As the present work shows, AGGAGG is indeed significantly underrepresented in the two non-zero frames (Fig. 1C and D). The corresponding data from zero frame occurrences of AGG_AGG is insignificant with respect to this point, since the numbers are so small.

**Twenty-five percent of the tandem AGA codons occur early in genes and likely affect ribosome loading.** The presence of rare codons close to the translational start site can have a strong negative effect on gene expression. Ribosomal stalling at early rare codons can lead to mRNA release from ribosomes (22) and termination of translation. Interestingly, this phenomenon is utilized for regulation of expression by a substantial number of genes in *E. coli*. Prior work showed that the rare codon AGA or AGG preferentially occurs early in a group of essential genes with a variety of functions (14, 15). Tandem rare codons near the 5′ end of a coding sequence confer an even stronger inhibitory effect on translation (15, 24, 49). It is only a small extension to now report that 11 of the 42 (26%) AGA_AGA tandem codons occur within the first 25 codons (Fig. 3 shows the distribution of the AGA_AGA positions within the genes). Intriguingly, half of these genes have a membrane-associated function. In one of the genes, *yhiU*, which encodes a putative membrane protein, AGA occurs as the triplet AGA_AGA_AGA.

**Other features relevant to AGA_AGA occurrences.** There are several features which may be relevant to occurrences of AGA_AGA. Suboptimal aminoacylation levels of sparse tRNAs (which decode rare codons) are a sensitive indicator of amino acid starvation. Starved cells need to preferentially synthesize essential proteins, and it is likely advantageous to abort the synthesis of less important proteins. Transfer mRNA (tmRNA)-mediated tagging that occurs at ribosomes stalled at rare codons may permit a redirection of ribosomes to more important tasks. tmRNA-mediated tagging occurs at tandem AGA codons, including in the *yjgR* gene as shown in Table 2, and may involve mRNA cleavage (27, 28, 48, 59). Consistent

with the redirection possibility, none of the genes with AGA_AGA are highly expressed (32), and none are known to be important under starvation conditions (Table 2). The occurrence of AGA_AGA early in genes (discussed above) is also preferable since cells will not waste energy synthesizing long peptides destined for degradation. A certain level of +1 frameshifting at AGA_AGA codons, whose level is dependent on sequence context, will partially circumvent tmRNA-mediated aborted synthesis. In most locations, such circumvention may be disadvantageous (exceptions may occur when AGA_AGA is positioned near the end of the gene or the 3′ ends of sequences encoding domains that are functional on their own).

Occurrences of AGA_AGA may also partially reflect the evolutionary history, by lateral transfer or direct descent, from an ancestor in which AGA was not shift prone due to the presence of multiple genes encoding its cognate tRNA. AGA codons occur at high frequency in toxin-encoding genes of *E. coli* O157:H7 (31), and the cognate $tRNA^{Arg}$ (anticodon UCU) occurs in eight copies in this bacterium (as annotated in the genome NC_002695). Therefore, AGA_AGA is not shift prone in O157:H7 and is not avoided. There are 62 genes containing AGA_AGA in O157:H7, and 28 of them have AGA_AGA-containing homologues in K-12. It could be that some of these 28 genes came to K-12 by horizontal gene transfer from O157:H7. Since none of these genes belong to the subset of highly expressed genes, evolutionary pressure is probably not strong enough to eliminate these shift-prone sequences from the K-12 genome. Furthermore, nearly 30% of the AGA_AGA occurrences in the *E. coli* K-12 genome are in prophages and mobile elements (Table 2).

**Perspective.** Any generalizations from the lack of evidence for utilization of frameshifting at tandem rare arginine codons need to be treated cautiously. Though no examples have yet been found, future work may reveal utilization of frameshifting at codons that become shift prone due to the limitation of the corresponding aminoacyl-tRNA following amino acid starvation. The penetrating work of Gallant and colleagues (19) clearly points out the possibilities. In conclusion, the present and a previous work (25) show that an error frameshift event(s), at levels considerably over the background rate, occurs during the translation of at least 2.5% of *E. coli* genes (see Materials and Methods). A number of other shift-prone sequences are known to occur in *E. coli*, and more are likely to be identified. It may well be that future studies will show that there is no strong selection against shift-prone sequences in general in ORFs that are not highly expressed.

## ACKNOWLEDGMENTS

We thank Jon Gallant for providing strains CP78 and CP79, Christopher Burge for providing the CodonShuffle and the DicodonShuffle programs, Emanuel Goldman and an anonymous referee for their constructive comments, and the *E. coli* Genetic Stock Center for providing strain BW27786 and the pLA2 and pINT-ts vectors.

This work was supported by grants from the DOE to R.F.G. and from the NIH (GM48152) to J.F.A., who was personally supported by Science Foundation Ireland at the final stage of this work.

## ADDENDUM IN PROOF

A recent publication (R. Leipuviene and G. R. Björk, RNA **11:**796–807, 2005) demonstrates that the tRNA cognate for at least one particular codon 5′ of a single AGA shifts to the +1 frame when there is either diminished quantity or functionality of the AGA-decoding tRNA. The degree of reduction of in-frame decoding of a single AGA needed before defects of cell division or other attributes become apparent (Leipuviene and Björk; K. Sakamoto, S. Ishimaru, T. Kobayashi, J. R. Walker, and S. Yokoyama, J. Bacteriol. **186:**5899–5905, 2004) is pertinent to the decoding of tandem AGA_AGA (or AGG_AGG), as discussed elsewhere (Leipuviene and Björk).

## REFERENCES

1. **Atkins, J. F., D. Elseviers, and L. Gorini.** 1972. Low activity of β-galactosidase in frameshift mutants of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **69:**1192–1195.
2. **Atkins, J. F., R. F. Gesteland, B. R. Reid, and C. W. Anderson.** 1979. Normal tRNAs promote ribosomal frameshifting. Cell **18:**1119–1131.
3. **Atkins, J. F., R. B. Weiss, and R. F. Gesteland.** 1990. Ribosome gymnastics–degree of difficulty 9.5, style 10.0. Cell **62:**413–423.
4. **Atkins, J. F., P. V. Baranov, O. Fayet, A. J. Herr, M. T. Howard, I. P. Ivanov, S. Matsufuji, W. A. Miller, B. Moore, M. F. Prère, N. M. Wills, J. Zhou, and R. F. Gesteland.** 2001. Overriding standard decoding: implications of recoding for ribosome function and enrichment of gene expression. Cold Spring Harbor Symp. Quant. Biol. **66:**217–232.
5. **Barak, Z., J. Gallant, D. Lindsley, B. Kwieciszewki, and D. Heidel.** 1996. Enhanced ribosome frameshifting in stationary phase cells. J. Mol. Biol. **263:**140–148.
6. **Baranov, P. V., R. F. Gesteland, and J. F. Atkins.** 2004. P-site tRNA is a crucial initiator of ribosomal frameshifting. RNA **10:**221–230.
7. **Baranov, P. V., R. F. Gesteland, and J. F. Atkins.** 2002. Recoding: translational bifurcations in gene expression. Gene **286:**187–201.
8. **Baranov, P. V., R. F. Gesteland, and J. F. Atkins.** 2002. Release factor 2 frameshifting sites in different bacteria. EMBO Rep. **3:**373–377.
9. **Baranov, P. V., O. L. Gurvich, A. W. Hammer, R. F. Gesteland, and J. F. Atkins.** 2003. RECODE 2003. Nucleic Acids Res. **31:**87–89.
10. **Belcourt, M. F., and P. J. Farabaugh.** 1990. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell **62:**339–352.
11. **Blinkowa, A. L., and J. R. Walker.** 1990. Programmed ribosomal frameshifting generates the *Escherichia coli* DNA polymerase III gamma subunit from within the tau subunit reading frame. Nucleic Acids Res. **18:**1725–1729.
12. **Boycheva, S., G. Chkodrov, and I. Ivanov.** 2003. Codon pairs in the genome of *Escherichia coli*. Bioinformatics **19:**987–998.
13. **Byström, A. S., K. J. Hjalmarsson, P. M. Wikström, and G. R. Björk.** 1983. The nucleotide sequence of an *Escherichia coli* operon containing genes for the tRNA(m1G)methyltransferase, the ribosomal proteins S16 and L19 and a 21-K polypeptide. EMBO J. **2:**899–905.
14. **Chen, G.-F., and M. Inouye.** 1990. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. Nucleic Acids Res. **18:**1465–1473.
15. **Chen, G. T., and M. Inouye.** 1994. Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. Genes Dev. **8:**2641–2652.
16. **Eguchi, Y., T. Oshima, H. Mori, R. Aono, K. Yamamoto, A. Ishihama, and R. Utsumi.** 2003. Transcriptional regulation of drug efflux genes by EvgAS, a two-component system in *Escherichia coli*. Microbiology **149:**2819–2828.
17. **Fiil, N., and J. D. Friesen.** 1968. Isolation of "relaxed" mutants of *Escherichia coli*. J. Bacteriol. **95:**729–731.
18. **Flower, A. M., and C. S. McHenry.** 1990. The gamma subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. Proc. Natl. Acad. Sci. USA **87:**3713–3717.
19. **Gallant, J., D. Lindsley, and J. Masucci.** 2000. The unbearable lightness of peptidyl-tRNA, p. 385–396. *In* R. A. Garrett, S. R. Douthwaite, A. Liljas, A. T. Matheson, P. B. Moore, and H. F. Noller (ed.), The ribosome: structure, function, antibiotics, and cellular interactions. ASM Press, Washington, D.C.
20. **Gallant, J., P. Bonthuis, D. Lindsley, J. Cabellon, G. Gill, K. Heaton, B. Kelley-Clarke, L. MacDonald, S. Mercer, H. Vu, and A. Worsley.** 2004. On the role of the starved codon and the takeoff site in ribosome bypassing in *Escherichia coli*. J. Mol. Biol. **342:**713–724.
21. **Gallant, J. A., and D. Foley.** 1980. On the causes and prevention of mistranslation, p. 615–638. *In* G. Chambliss, G. R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (ed.), Ribosomes, structure, function and genetics. University Park Press, Baltimore, Md.
22. **Gao, W., S. Tyagi, F. R. Kramer, and E. Goldman.** 1997. Messenger RNA release from ribosomes during 5′-translational blockage by consecutive low-usage arginine but not leucine codons in *Escherichia coli*. Mol. Microbiol. **25:**707–716.
23. **Gao, X., E. R. Havecker, P. V. Baranov, J. F. Atkins, and D. F. Voytas.** 2003.

Translational recoding signals between *gag* and *pol* in diverse LTR retrotransposons. RNA **9:**1422–1430.
24. **Goldman, E., A. H. Rosenberg, G. Zubay, and F. W. Studier.** 1995. Consecutive low-usage leucine codons block translation only when near the 5′ end of a message in *Escherichia coli*. J. Mol. Biol. **245:**467–473.
25. **Gurvich, O. L., P. V. Baranov, J. Zhou, A. W. Hammer, R. F. Gesteland, and J. F. Atkins.** 2003. Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. EMBO J. **22:**5941–5950.
26. **Haldimann, A., and B. L. Wanner.** 2001. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J. Bacteriol. **183:**6384–6393.
27. **Hayes, C. S., B. Bose, and R. T. Sauer.** 2002. Stop codons preceded by rare arginine codons are efficient determinants of SsrA tagging in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **99:**3440–3445.
28. **Hayes, C. S., and R. T. Sauer.** 2003. Cleavage of the A-site mRNA codon during ribosome pausing provides a mechanism for translational quality control. Mol. Cell **12:**903–911.
29. **Herr, A. J., C. C. Nelson, N. M. Wills, R. F. Gesteland, and J. F. Atkins.** 2001. Analysis of the roles of tRNA structure, ribosomal protein L9, and the bacteriophage T4 gene 60 bypassing signals during ribosome slippage on mRNA. J. Mol. Biol. **309:**1029–1048.
30. **Ivanov, I. P., R. F. Gesteland, and J. F. Atkins.** 2000. Antizyme expression: a subversion of triplet decoding, which is remarkably conserved by evolution, is a sensor for an autoregulatory circuit. Nucleic Acids Res. **28:**3185–3196.
31. **Kanjo, N., and H. Inokuchi.** 1999. Genes for tRNA(Arg) located in the upstream region of the Shiga toxin II operon in enterohemorrhagic *Escherichia coli* O157:H7. DNA Res. **6:**71–73.
32. **Karlin, S., J. Mrázek, A. Campbell, and D. Kaiser.** 2001. Characterizations of highly expressed genes of four fast-growing bacteria. J. Bacteriol. **183:** 5025–5040.
33. **Katz, L., and C. B. Burge.** 2003. Widespread selection for local RNA secondary structure in coding regions of bacterial genes. Genome Res. **13:**2042–2051.
34. **Khlebnikov, A., K. A. Datsenko, T. Skaug, B. L. Wanner, and J. D. Keasling.** 2001. Homogeneous expression of the $P_{BAD}$ promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology **147:**3241–3247.
35. **Kurland, C. G., D. Hughes, and M. Ehrenberg.** 1996. Limitations of translational accuracy, p. 979–1004. *In* F. C. Niedhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Resnikoff, M. Riley, M. Schaechter and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, D.C.
36. **Larsen, B., N. M. Wills, R. F. Gesteland, and J. F. Atkins.** 1994. rRNA-mRNA base pairing stimulates a programmed −1 ribosomal frameshift. J. Bacteriol. **176:**6842–6851.
37. **Larsen, B., N. M. Wills, C. Nelson, J. F. Atkins, and R. F. Gesteland.** 2000. Nonlinearity in genetic decoding: homologous DNA replicase genes use alternatives of transcriptional slippage or translational frameshifting. Proc. Natl. Acad. Sci. USA **97:**1683–1688.
38. **Lindsley, D., and J. Gallant.** 1993. On the directional specificity of ribosome frameshifting at a "hungry" codon. Proc. Natl. Acad. Sci. USA **90:**5469–5473.
39. **Masucci, J. P., J. Gallant, D. Lindsley, and J. Atkinson.** 2002. Influence of the *relA* gene on ribosome frameshifting. Mol. Genet. Genomics **268:**81–86.
40. **Masuda, N., and G. M. Church.** 2002. *Escherichia coli* gene expression responsive to levels of the response regulator EvgA. J. Bacteriol. **184:**6225–6234.
41. **Matsufuji, S., T. Matsufuji, Y. Miyazaki, Y. Murakami, J. F. Atkins, R. F. Gesteland, and S. Hayashi.** 1995. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. Cell **80:**51–60.
42. **Miller, J. H.** 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
43. **Namy, O., J. P. Rousset, S. Napthine, and I. Brierley.** 2004. Reprogrammed genetic decoding in cellular gene expression. Mol. Cell **13:**157–168.
44. **Neidhardt, F. C., P. L. Bloch, and D. F. Smith.** 1974. Culture medium for enterobacteria. J. Bacteriol. **119:**736–747.
45. **O'Connor, M.** 2002. Imbalance of $tRNA^{Pro}$ isoacceptors induces +1 frameshifting at near-cognate codons. Nucleic Acids Res. **30:**759–765.
46. **Pande, S., A. Vimaladithan, H. Zhao, and P. J. Farabaugh.** 1995. Pulling the ribosome out of frame by +1 at a programmed frameshift site by cognate binding of aminoacyl-tRNA. Mol. Cell. Biol. **15:**298–304.
47. **Régnier, P., M. Grunberg-Manago, and C. Portier.** 1987. Nucleotide sequence of the *pnp* gene of *Escherichia coli* encoding polynucleotide phosphorylase. Homology of the primary structure of the protein with the RNA-binding domain of ribosomal protein S1. J. Biol. Chem. **262:**63–68.
48. **Roche, E. D., and R. T. Sauer.** 1999. SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. EMBO J. **18:**4579–4589.
49. **Rosenberg, A. H., E. Goldman, J. J. Dunn, F. W. Studier, and G. Zubay.** 1993. Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. J. Bacteriol. **175:**716–722.
50. **Ruusala, T., D. Andersson, M. Ehrenberg, and C. G. Kurland.** 1984. Hyperaccurate ribosomes inhibit growth. EMBO J. **3:**2575–2580.
51. **Seffens, W., and D. Digby.** 1999. mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. Nucleic Acids Res. **27:**1578–1584.
52. **Shah, A. A., M. C. Giddings, J. B. Parvaz, R. F. Gesteland, J. F. Atkins, and I. P. Ivanov.** 2002. Computational identification of putative programmed translational frameshift sites. Bioinformatics **18:**1046–1053.
53. **Sipley, J., and E. Goldman.** 1993. Increased ribosomal accuracy increases a programmed translational frameshift in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **90:**2315–2319.
54. **Spanjaard, R. A., K. Chen, J. R. Walker, and J. van Duin.** 1990. Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to *argU* tRNA and T4 $tRNA^{Arg}$. Nucleic Acids Res. **18:**5031–5036.
55. **Spanjaard, R. A., and J. van Duin.** 1988. Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. Proc. Natl. Acad. Sci. USA **85:**7967–79671.
56. **Stahl, G., S. Ben Salem, Z. Li, G. McCarty, A. Raman, M. Shah, and P. J. Farabaugh.** 2001. Programmed +1 translational frameshifting in the yeast *Saccharomyces cerevisiae* results from disruption of translational error correction. Cold Spring Harbor Symp. Quant. Biol. **66:**249–258.
57. **Stahl, G., S. N. Ben Salem, L. Chen, B. Zhao, and P. J. Farabaugh.** 2004. Translational accuracy during exponential, postdiauxic, and stationary growth phases in *Saccharomyces cerevisiae*. Eukaryot. Cell **3:**331–338.
58. **Sundararajan, A., W. A. Michaud, Q. Qian, G. Stahl, and P. J. Farabaugh.** 1999. Near-cognate peptidyl-tRNAs promote +1 programmed translational frameshifting in yeast. Mol. Cell **4:**1005–1015.
59. **Sunohara, T., K. Jojima, Y. Yamamoto, T. Inada, and H. Aiba.** 2004. Nascent-peptide-mediated ribosome stalling at a stop codon induces mRNA cleavage resulting in nonstop mRNA that is recognized by tmRNA. RNA **10:**378–386.
60. **Tanabe, H., K. Yamasak, M. Furue, K. Yamamoto, A. Katoh, M. Yamamoto, S. Yoshioka, H. Tagami, H. A. Aiba, and R. Utsumi.** 1997. Growth phase-dependent transcription of *emrKY*, a homolog of multidrug efflux *emrAB* genes of *Escherichia coli*, is induced by tetracycline. J. Gen. Appl. Microbiol. **43:**257–263.
61. **Tsuchihashi, Z., and A. Kornberg.** 1990. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. Proc. Natl. Acad. Sci. USA **87:**2516–2520.
62. **Urbonavi[caron]cius, J., Q. Qian, J. M. B. Durand, T. G. Hagervall, and G. R. Björk.** 2001. Improvement of reading frame maintenance is a common function for several tRNA modifications. EMBO J. **20:**4863–4873.
63. **Weiss, R. B., D. M. Dunn, J. F. Atkins, and R. F. Gesteland.** 1987. Slippery runs, shifty stops, backward steps, and forward hops: −2, −1, +1, +2, +5, and +6 ribosomal frameshifting. Cold Spring Harbor Symp. Quant. Biol. **52:**687–693.
64. **Weiss, R. B., D. M. Dunn, A. E. Dahlberg, J. F. Atkins, and R. F. Gesteland.** 1988. Reading frame switch caused by base-pair formation between the 3′ end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. EMBO J. **7:**1503–1507.
65. **Wenthzel, A. M., M. Stancek, and L. A. Isaksson.** 1998. Growth phase dependent stop codon readthrough and shift of translation reading frame in *Escherichia coli*. FEBS Lett. **421:**237–242.
66. **Wills, N. M., J. A. Ingram, R. F. Gesteland, and J. F. Atkins.** 1997. Reported translational bypass in a *trpR′-lacZ′* fusion is accounted for by unusual initiation and +1 frameshifting. J. Mol. Biol. **271:**491–498.
67. **Xu, J., R. W. Hendrix, and R. L. Duda.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol. Cell **16:**11–21.
68. **Zahn, K.** 1996. Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. J. Bacteriol. **178:**2926–2933.
69. **Zimmer, M., E. Sattelberger, R. B. Inman, R. Calendar, and M. J. Loessner.** 2003. Genome and proteome of *Listeria monocytogenes* phage PSA: an unusual case for programmed +1 translational frameshifting in structural protein synthesis. Mol. Microbiol. **50:**303–317.
70. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406–3415.